УДК 004.318

В.П. Бурдаев
Харьковский национальный экономический университет, Украина
Украина, 61001, г. Харьков, пр. Ленина, 9а

Фильтрация базы знаний

V.P. Burdaev
Kharkov National Economic University, Ukraine
Ukraine, 61001, c. Kharkov, Lenin's avenue, 9a

Knowledge Base Filtration

В.П. Бурдаєв
Харківський національний економічний університет, Україна
Україна, 61001, м. Харків, пр. Леніна, 9а,

Фільтрація бази знань

Рассматривается модель иерархической функциональной системы динамической предметной области. Предлагаемая модель состоит из иерархии классов предметной области, связей между ними (правил вывода), которые действуют в рамках этой модели. Рассматривается задача фильтрации базы знаний с позиций адаптации ее правил к динамической предметной области. Предложен, реализован и исследуется механизм интерпретации модели иерархической функциональной системы в условиях динамического изменения ее параметров (базового класса, связей между классами и взаимодействия объектов классов).
Ключевые слова: фильтрация базы знаний, открытая система, иерархическая функциональная система, возмущения правил базы знаний, динамическая онтология.

The model of hierarchical functional system of dynamic data domain is considered. The offered model consists of hierarchy of classes of data domain, communications between them (conclusion rules) which operate within the limits of this model. The problem of a filtration of the knowledge base from positions of adaptation of its rules to dynamic data domain is considered. The mechanism of interpretation of model of hierarchical functional system in the conditions of dynamic change of its parametres (a base class, communications between classes and interactions of objects of classes) is offered, realised and is investigated.
Keywords: the knowledge base filtration, open system, hierarchical functional system, indignations of rules of the knowledge base, dynamic онтология.

Розглядається модель ієрархічної функціональної системи динамічної предметної області. Пропонована модель складається з ієрархії класів предметної області, зв’язків між ними (правил виводу), які діють у рамках цієї моделі. Розглядається завдання фільтрації бази знань із позицій адаптації її правил до динамічної предметної області. Запропоновано, реалізований і досліджується механізм інтерпретації моделі ієрархічної функціональної системи в умовах динамічної зміни її параметрів (базового класу, зв'язків між класами й взаємодії об’єктів класів).
Ключові слова: фільтрація бази знань, відкрита система, ієрархічна функціональна система, збурювання правил бази знань, динамічна онтологія.

Распределенные интеллектуальные системы – это качественно новые технологии, особенностями которых являются моделирование функциональных систем, использование динамически развивающейся онтологии предметной области, мультиагентный подбор адаптивной стратегии принятия решения.

Практика показывает, что главной в теории сложных систем становится проблема принятия решений при наличии многих целей. Динамической следует считать систему, которая обладает такими свойствами, как связность, сложность, устойчивость и цели поведения, которой слабо формализованы.

Современные информационные предметные области обладают динамической структурой, как например, сети Интернет, предсказание аварийных и чрезвычайных ситуаций, распределенное обучение и т.д. Их особенностями являются: наличие огромного числа автономных сущностей со своими конкретными подцелями (автономность). Сущности подвержены воздействиям внешней среды (открытость), взаимодействуют между собой (распределенность). Базы знаний сущностей уникальны (локальность) и образуют иерархические коалиции (иерархия уровней сущностей). Для построения моделей баз знаний таких предметных областей используют, например, самоорганизующие открытые многоагентные системы [1].

Для изучения глобальных свойств динамических процессов в предметных областях применяют математические структуры из дифференциальной топологии – многообразия, которые интерпретируются как «пространство состояний», а динамический процесс моделируют с помощью дифференцируемых отображений (диффеоморфизмов или гладких потоков). Техника глобального анализа во многом опирается на предположение о компактности многообразия без края, что не выполняется для информационных потоков в предметных областях. Поэтому использовать классические определения динамической системы, которая рассматривается как однопараметрическая группа преобразований топологического пространства M, удовлетворяющих условиям:

1) $f(x,0) = x$ ($f : M \times T \rightarrow M$ – непрерывное отображение, $x \in M$, 0 – нуль топологической группы T),

2) $f(f(x,t),s) = f(x, t + s)$ ($x \in M$, $t, s \in T$),

не приходится.

Поэтому под моделью динамической предметной областью понимаем функциональную систему, которая характеризуется некоторым состоянием в данный момент времени и некоторым законом, который описывает изменение состояние с течением времени. А именно состояние функциональной системы описывается фактами предметной области, а изменение состояния описывается правилами принятия решения. Один из подходов построения динамических интеллектуальных систем, основанных на правилах, рассмотрен работе в [2].

В работе [3] построена математическая модель динамической предметной области в виде иерархической функциональной системы (ФС), в которой база знаний ассоциируется с цепочкой расслоений баз знаний, т.е. представляет собой сечение цепочки расслоений баз знаний.

Модели открытых (живых) систем описываются системами неавтономных дифференциальных уравнений. Для изучения топологических свойств, структурной устойчивости и управления такими системами применяются расширения динамических систем на основе глобального анализа.

Пусть задано гладкое расширение $p:(M, R, \pi) \rightarrow (B, R, \rho)$, разлагающееся в цепочку расширений $p_i:(X_i, R, \pi_i) \rightarrow (X_{i+1}, R, \pi_{i+1})$ ($i=0,1,\ldots,n-1$), т.е. $(X_0, R, \pi_0)=(M, R, \pi)$, $(X_n, R, \pi_n)= (B, R, \rho)$ и $p_{n-1} \circ \ \circ p_1 \circ p_0=p$.

Задание этой цепочки можно интерпретировать как выделение в системе (M, R, π) иерархии уровней (X_i, R, π_i) ($i=0,1,\ldots,n-1$).

Другими словами, модель иерархической ФС представлена сечениями следующей коммутативной диаграммы $p \circ \pi = \rho \circ p$ на рис. 1.

При этом каждое сечение цепочки расслоений (ФС) имеет вид орграфа изучаемого целевого объекта при фиксированном $b \in B$, другими словами, состояние ФС описы-

вается в заданный момент времени. Изменение сечения прообразов цепочки расслоения базы знаний (иерархической ФС) вовремя обработки их агентом вывода в динамике представлено на рис. 2.

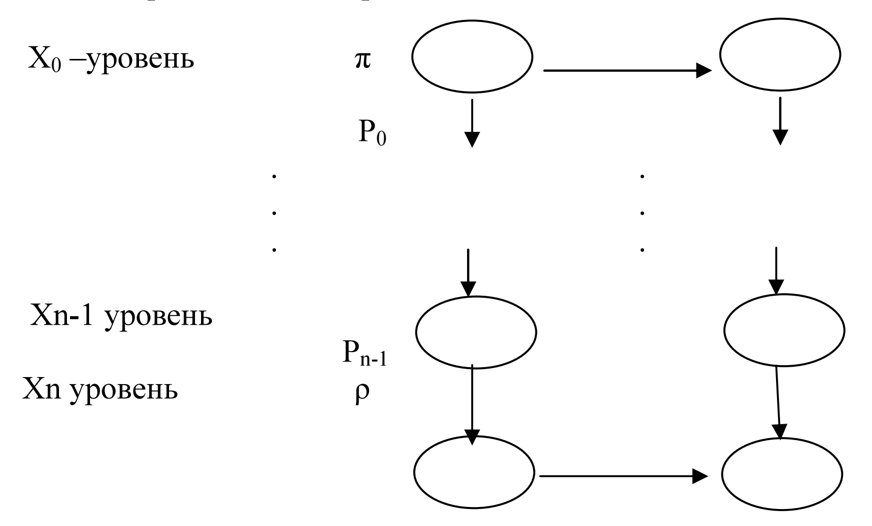


Рисунок 1 – Модель иерархической ФС

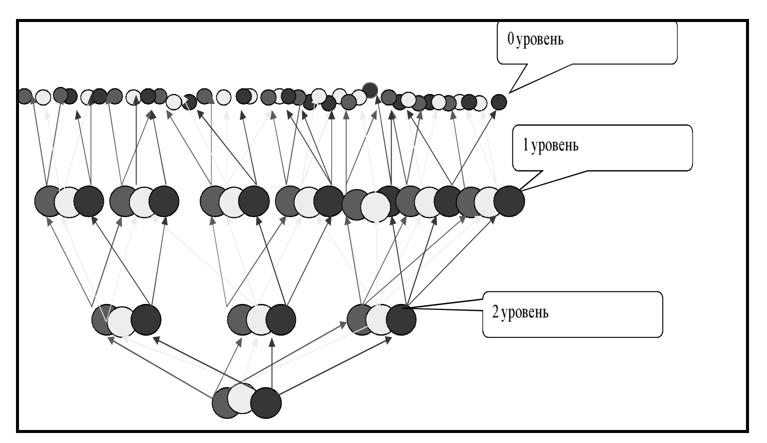


Рисунок 2 – Динамическое изменение сечения расслоения базы знаний

ФС характеризуется следующими свойствами:

– связностью – цепочка расслоений базы знаний;

– сложностью – иерархия уровней локальных баз знаний;

– устойчивостью (динамическое поведение системы) – структура сечения цепочки расслоений не меняется при возмущениях локальных баз знаний. Другими словами, осуществляется только изменение эвристик локальных баз знаний цепочки расслоений, а база расслоения, которая интерпретируется как внешняя среда, остается неизменной.

Связность ФС выражается в фильтрации базы знаний. Пусть B_i – это локальная база знаний, то есть содержит правила продукций для определения подцели Gi , которая находится на i-м уровне в иерархии ФС.

Фильтрация базы знаний – это конечная система локальных баз знаний B_i

$$B_0 <= B_1 <= \ldots <= B_k,$$

частично-упорядочных (<=) следующим образом: консеквент каждого правила из B_i содержится в антецеденте правила из B_j.

Чтобы построить фильтрацию базы знаний, достаточно указать цепочку правил ФС для достижения основной цели. Затем с помощью рекурсивного алгоритма строится, другие правила для локальных баз знаний ФС путем генерирования правил вовремя консультации с экспертом. Эксперт анализирует правило, предоставленное агентом вывода ФС, и может поместить созданные правила в локальную базу знаний на соответствующем уровне иерархии ФС или запретить его использование, или разрешить его использование на определенное время. Таким образом, с помощью алгоритма фильтрации базы осуществляется как пополнение локальных баз знаний, так и их адаптация к предметной области. Количество локальных баз знаний соответствует уровням иерархической ФС предметной области.

Сложность ФС характеризует основное свойство иерархической системы: несмотря на наличие ошибок в локальных базах вся ФС в целом функционирует нормально. Например, если на каком-то уровне иерархии не нашлось соответствующего правила для принятия решения, то ситуация автоматически агентом вывода заносится во фрейм и далее производится переход к следующему уровню; другими словами консультация не прерывается. Кроме того, агент вывода формирует правило продукции на эту ситуацию ФС и запоминает в базе знаний. После консультации эксперт может проанализировать ситуацию, которая привела к появлению нового правила принятия решения.

Заметим, что размерность ФС (множество состояний или множество правил принятия решений) не означает большую ее сложность и наоборот.

Интуитивно под структурной устойчивостью функциональной системы ФС будем понимать существование некоторой «окрестности» орграфа ФС, в которой орграфы «близких» к ней систем в некотором смысле «близки» исходному (рис. 2).

Таким образом, под динамической системой будем понимать именно иерархическую функциональную систему предметной области, в которой правила принятия решения не разбросаны, а представляют собой набор отфильтрованных локальных баз знаний, которые позволяют агенту ФС достичь локальной цели на каждом уровне иерархии и соответственно – глобальной цели.

Итак, модель предметной области рассматривается как ФС, в которой результат оказывает организующее влияние на все этапы формирования онтологии. Классы и связи между ними можно рассматривать как логическую конструкцию ФС.

Пусть B – множество атрибутов предметной области, V – множество их значений, $p: B \times V \to B$ – проекция, $V_b \equiv p^{-1}(b)$ – слой расслоения над точкой b, то есть множество значений атрибута b. В прямом произведении $B \times V$ определим отношение R, наделенное следующим свойством: для любого атрибута $b \in B$ существует значение $v \in V_b$, такое, что $R \equiv bRv$. Другими словами, отношение R определяет принадлежность значений из слоя V_b конкретному атрибуту, то есть задает факты предметной области.

Пусть Ф – множество фактов предметной области (ПрО) и $r : Ф \to Ф$ – правило (если <условие> то <действие>). Тогда объекты ПрО определяются как атрибуты, которые находятся в правой части правила и формируют факты только в том случае, если правило выполняется (<условие> – истинно). Среди объектов выделяется целевой, который является системоорганизующим. Затем выделяется другие объекты, которые располагаются в определенной иерархии в соответствии с их правилами и могут быть отображены с помощью орграфа.

Пусть система $F \subset B \times V$ параметризована некоторым множеством N в том смысле, что каждому $t \in N$ соответствует некоторое состояние работы системы, которое отображается орграфом. При этом следующее состояние системы определяется текущими фактами и правилами на соответствующих уровнях иерархии.

Реализация расслоения базы знаний в цепочку расслоений позволяет эффективно выполнить ряд операций:

1. Тестирования базы знаний на каждом уровне иерархической ФС (поиск непротиворечивости знаний и исследование на полноту базы знаний). Клонирование правил базы знаний.

2. Выполнить наглядное представление движения сечений цепочки расслоения базы знаний при эволюционном перемещении базового класса в базе расслоения. База расслоения – это так называемые медленные переменные, описывающие эволюцию главной цели. Слои цепочки расслоений – это быстрые переменные, описывающие состояния объектов, подцелей в дискретные моменты времени при эволюции главной цели.

3. Производить вертикальное возмущение в слоях цепочки расслоения для поиска скрытых знаний. Другими словами, в иерархической ФС происходит возмущение не всей базы знаний, а только в слоях. Таким образом, основная цель иерархической ФС не подвергается возмущениям внешней среды, а эволюционирует во времени для достижения главного результата.

4. Использовать агентов (классы в системе «КАРКАС» [5]) не только для управления консультацией на разных уровнях иерархической ФС, но и для объяснения результатов, получаемых в результате консультации.

5. Выполнить тестирование знаний пользователя по предметной области (возможно, использовать до десяти различных типов вопросов) и предоставить визуализацию тестирования в виде статистики, диаграмм и модифицированных «лиц Чернова». Для формирования теста формируется специфическая модель ФС.

6. Имеется возможность провести интеллектуальный анализ данных. Построена специальная модель ФС, которая позволяет подобрать алгоритм кластеризации, исходя из описаний структуры многомерных данных.

Другой подход к моделированию динамической предметной области — это использование темпоральных знаний [6].

По аналогии с моделированием открытых систем с помощью расширений динамических систем рассмотрим вертикальные возмущения правил базы знаний. Под возмущением можно понимать как их поглощение ФС, другими словами ФС адаптируется к внешней среде.

Пусть $A = \{a_1, a_2, ..., a_n\}$ – множество атрибутов предметной области, V_j – множество возможных значений $a_j \in A$ и $V = U_{a \in A} V_a$. Атрибуты могут быть измерены в разных шкалах (количественной, порядковой, качественной, смешанной).

Между атрибутом a_j и его значением v_j определим следующие операции:

1) $a_j = v_j$, $=$ – знак операции означивания;

2) $a_j < v_j$, $<$ – знак операции отношения ($>$, $>=$, $<=$);

3) $a_j \in [v_{j1}, ..., v_{jm}]$, $\in$ – знак операции принадлежности ([ ), ( ), ( ]).

Атрибут, его значение и операция между ними определяют высказывание.

Пусть $Q = \{q: A \text{->} V\}$ – множество высказываний.

На множестве Q определим две функции:

1. Функция $\mu : Q \to [0,\ldots,1]$ позволяет интерпретировать высказывание q с точки зрения его истинности.

2. Функция $\theta : Q \to [0,\ldots,1]$ позволяет интерпретировать высказывание q с точки зрения его значимости.

Тогда триплет $f = (q, \mu(q), \theta(q))$, где $q \in Q$, назовем фактом предметной области.

Пусть λ отображение из прямого произведения $Q \times [0,\ldots,1] \times [0,\ldots,1]$ в $Q \times [0,\ldots,1] \times [0,\ldots,1]$. Через $\Gamma = \{ \lambda \}$ множество всех возможных отображений таких, что $\lambda(f) \in Q \times [0,\ldots,1] \text{ x } [0,\ldots,1]$.

Определим состояние функциональной системы s как набор фактов, связанных между собой знаками конъюнкции, дизъюнкции или отрицанием. Например,

$$s = f_1 \;\&\; f_2 \;\; U \;\; (\neg\, f_3).$$

Обозначим через S = {s} множество всех состояний предметной области.

Пусть $R^* = \{r: S \rightarrow S\}$ – множество преобразований (правил) следующего вида:

r : ЕСЛИ <АНТЕЦЕДЕНТ> ТО <КОНСЕКВЕНТ>,

где <АНТЕЦЕДЕНТ> (условие – ситуация s_1), <КОНСЕКВЕНТ> (вывод, следствие – ситуация s_2).

Если ситуация s_1 в правиле принимает значение истина, тогда ситуация s_2 получает значение истина и добавляется к S. Другими словами, если ситуация $s_1 \in S$ в правиле принимает значение истина, тогда существует отображение q_j: A -> V, формирующее ситуацию s_2

$$s_2 = q_1 * q_2 \quad * q_n$$ и * знак операций & , U ,¬.

Если ситуация s_1 в правиле принимает значение ложь, тогда ситуация s_2 не добавляется к S. Например, ЕСЛИ s ТО q_4, Здесь s – ситуация (3), q_4 – некоторое высказывание. Результатом выполнения правила будет создание нового факта:

$$f_4 = (q_4 , \mu(q_4), \theta(q_4)),$$

$$\mu(q_4) = k * \max [\min(\mu (q_1), \mu (q_2)), (1 - \mu (q_3))],$$

$$\theta(q_4) = k * \max [\min(\theta(q_1), \theta(q_2)), (1 - \theta(q_3))].$$

Пусть триплет $f_{t=0} = (q, \mu(q), \theta(q))$, где $q \in Q$ – факт предметной области, образованный в момент времени $t_0 = 0$. Если в момент времени $t = t_1$ выбирается другое значение из домена атрибута, обозначим $f_{t1,}$ и во всех правилах БЗ происходит замена факта f_{t0} на $f_{t1,}$ то такое действие над БЗ назовем возмущением ФС в момент времени $t = t_1$. Заметим, что возмущение узкое (вертикальное), поскольку при нем логическая структура ФС (орграф, другими словами, сечение цепочки расслоении, определяющее ФС), последовательность расположения правил в БЗ не меняется.

Таким образом, получаем состояние ФС в момент времени $t = t_1$. Это состояние можно сохранить, записав его в отдельный файл, провести тестирование и проанализировать ход консультации на предмет содержательности БЗ в момент времени $t = t_1$. Если эксперт считает, что состояние БЗ в момент $t = t_1$ важно, то он может рекомендовать пользователю БЗ для эксплуатации. Если нет, то можно вызвать возмущение ФС в момент времени $t = t_2$ и повторить предыдущие шаги. Затем можно активизировать другой атрибут и проанализировать изменения БЗ. В результате таких специальных (вертикальных) возмущений БЗ эволюционирует, и эксперт может найти когнитивную ситуацию, которая не была им обнаружена при составлении БЗ. Такие возмущения БЗ используются на стадии ее отладки.

Литература

1. Городецкий В.И. Самоорганизация и многоагентные системы. Модели многоагентной самоорганизации. / В.И. Городецкий // Известия РАН «Теория и системы управления». – 2012. – № 2. – С. 92-120.
2. Осипов Г. С. Динамические интеллектуальные системы. / Г.С. Осипов // Искусственный интеллект и принятие решений. – 2008. – № 1. – С. 47-54.
3. Бурдаев В. П. Модель функциональной системы динамической предметной области. / В.П.Бурдаев // Искусственный интеллект – 2011. – № 3. – С.439-448.
4. Бурдаєв В. П. Моделі баз знань. / В.П. Бурдаев – Харків : Вид. ХНЕУ, 2010. – 300 с.

5. Режим доступа : http://www.it-karkas.com.ua − компьютерная система "КАРКАС".
6. Еремеев А.П. Логика ветвящегося времени и ее применение в интеллектуальных системах поддержки принятия решений. // Десятая национальная конференция по искусственному интеллекту с междунар. участием КИИ-2006 : труды конф. в 3 т. – М.: Физматлит, 2006. – Т.3.

Literatura

1. Gorodetsky V. I. Self-organising and многоагентные systems. Models многоагентной self-organising. / V.I.Gorodetsky // News of the Russian Academy of Sciences "the Theory and management systems", 2012. – № 2. – P. 92-120.
2. Osipov G. S. Dynamic intellectual systems. / G.S.Osipov // Machine intelligence and decision-making. - 2008. – № 1. – P. 47 - 54.
3. Burdaev V.P. Model of functional system of dynamic data domain. / V.P. Burdaev // Machine intelligence – 2011. – №3. – P.439-448.
4. Burdaev V.P. Model_ of bases знань. / V.P. Burdaev - Kharkiv: KhNEU, 2010. – 300 P.
5. http://www.it-karkas.com.ua − computer system "KARKAS".
6. Yeremeyev А.П. Logic time and its application in intellectual systems of support of decision-making. // Tenth national conference on artificial intelligence with the international participation KII-2006: Works conf. In 3 volumes. – M: Fizмatlit, 2006. Т.3.

RESUME

V.P.Burdaev

Knowledge Base Filtration

In given article the concept of a filtration of the knowledge base realised in the form of hierarchical functional system which has shown the efficiency by working out онтологий in various information dynamic data domains is considered: medicine, economy, mobile communication and кластерный the analysis of the multidimensional data.

The computer model hierarchical ФС which is the formalized reflexion of data domain in the form of hierarchical structure of a set of managing directors a component (agents) which co-operate among themselves for overall objective achievement is realised.

Technical realisation of computer model of hierarchical functional system is executed in tool to the environment "KARKAS" which is intended for construction of computer models of the multiconnected systems on the basis of stratifications of knowledge bases of different function: economic, ecological, information etc.

Статья поступила в редакцию23.04.2013.